UNCLASSIFIED

# AD-A205 924

(2)

DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

ELECTE
MAR 1 0 1989
H

| | |
|---|---|
| 1a. REPORT SECURITY CLASSIFICATION<br>UNCLASSIFIED | 1b. RESTRICTIVE MARKINGS |
| 2a. SECURITY CLASSIFICATION AUTHORITY | 3. DISTRIBUTION/AVAILABILITY OF REPORT<br>Approved for Public Release;<br>Distribution Unlimited |
| 2b. DECLASSIFICATION/DOWNGRADING SCHEDULE | |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S) | 5. MONITORING ORGANIZATION REPORT NUMBER(S)<br>AFOSR-TR. 89-0286 |
| 6a. NAME OF PERFORMING ORGANIZATION<br>WEIDLINGER ASSOCIATES<br>6b. OFFICE SYMBOL (If applicable) | 7a. NAME OF MONITORING ORGANIZATION<br>AFOSR/NA |
| 6c. ADDRESS (City, State, and ZIP Code)<br>333 Seventh Avenue<br>New York, NY 10001 | 7b. ADDRESS (City, State, and ZIP Code)<br>Bldg. 410<br>Bolling AFB, DC 20332-6448 |
| 8a. NAME OF FUNDING/SPONSORING ORGANIZATION<br>AFOSR<br>8b. OFFICE SYMBOL (If applicable)<br>NA | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br>F49620-85-C-0045DEF |
| 8c. ADDRESS (City, State, and ZIP Code)<br>Bldg. 410<br>Bolling AFB, DC 20332-6448 | 10. SOURCE OF FUNDING NUMBERS<br>PROGRAM ELEMENT NO. 6.1102F; PROJECT NO. 2302; TASK NO C1; WORK UNIT ACCESSION NO. |

11. TITLE (Include Security Classification) (U) UNCERTAINTIES IN SOIL CONSTITUTIVE BEHAVIOR

12. PERSONAL AUTHOR(S) H. BENAROYA

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT (Year, Month, Day) | 15. PAGE COUNT |
|---|---|---|---|
| FINAL | FROM 3/1/85 TO 9/30/88 | 89/2/1 | 33 |

16. SUPPLEMENTARY NOTATION

| 17. COSATI CODES | | | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Phenomenological, Markov transition probability; Stochastic matrices; Convergence properties |
| | | | |
| | | | |

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

A Markov chain phenomenological framework is used to model soil constitutive behavior accounting for uncertainties. The key in a Markov chain model is its transition probability (stochastic) matrix. Two lines of study have been pursued: explore the properties of stochastic matrices with the purpose of explaining different classes of behavior according to the mathematical properties of these transition matrices; identify a relatively simple technique to estimate the transition probabilities from available experimental data. These were parallel efforts. The first is greatly enhanced by the success of the second.

| 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT | 21. ABSTRACT SECURITY CLASSIFICATION |
|---|---|
| ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS | UNCLASSIFIED |

| 22a. NAME OF RESPONSIBLE INDIVIDUAL | 22b. TELEPHONE (Include Area Code) | 22c. OFFICE SYMBOL |
|---|---|---|
| Major Steven C. Boyce | (202) 767-6963 | AFOSR/NA |

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

# UNCERTAINTIES IN SOIL CONSTITUTIVE BEHAVIOR

**Final Report: 3/1/85 – 9/30/88**

H Benaroya
Applied Science Division
Weidlinger Associates
333 Seventh Avenue
New York, New York 10001

ACKNOWLEDGMENTS

Prepared under Air Force Office of Scientific Research Contract No. F49620-85-C-0045DEF

Lawrence Hokanson, Spencer T Wu, and Steven C Boyce, of the Air Force Office of Scientific Research, are sincerely thanked for the financial support which has made this work possible.

19 October 1988
revised submission: 1 February 1989

## SUMMARY of FINAL REPORT

Our work begins with a brief review of important deterministic and probabilistic phenomenological soil modeling [1]. Behavioral ranges and associated problems are defined with respect to random excitation, material modeling, free–field behavior, and structure–media interaction. Next, a number of ideas are explored as possible frameworks for the phenomenological modeling of soil constitutive behavior, accounting for uncertainties [2], and a significant source of experimental data especially useful in our model development is identified [7]. Considering various possible approaches, it is concluded that a Markov framework provides the best approach since:

1. The fundamental axiom upon which all Markov theory rests is the following: the probability that a system occupies some state at time $t_1$ later than time $t$ depends on its disposition at time $t$ and not on any time earlier than $t$. This is analogous to the approach of (deterministic) classical mechanics where the evolution of a system is established given its present state. Thus, the Markov state transition matrix appears to be an ideal probabilistic counterpart to the transfer matrix concept used extensively in deterministic modeling, where this transition matrix and corresponding probability distribution of states couples the system evolutionary dynamics with system properties.
2. The need to know only the most recent state of the system is attractive also from the computational point of view. It would be possible to update the state of the system while retaining data from only one additional state. The complete history is not needed since it is effectively incorporated in this most recent information. This property also turns out to be useful in linking the method to experimental data.

The central thread through our efforts is the study of the Markov transition probability matrix. The main purpose is to explain different classes of dynamic behavior and constitutive properties in terms of the mathematical properties of transition matrices. These transition matrices are called stochastic matrices, and they have many properties; a key property from which all others follow is that row sums equal 1, that is, the probability of being in any of all possible states must equal 1 if the set of all possible states is all inclusive. The bulk of this work is reported in two reports [15,16].

As an aside to the focus of our efforts, ideas in fractal geometries as applied to the problem at hand drew our attention, and we put a few ideas of our own down on paper for future reference [17]. Given the opportunity and time, we may go back to these.

A parallel component of our efforts centered about linking the Markov transition probability matrix to experimental data. This is a key to using Markov models, and, ironically, a topic that does not appear to draw much attention in the applied literature. *How does the analyst/modeler take experimental data, extract in a rational way its essence, and build a transition probability matrix that reproduces that data in a probabilistic sense?* This was viewed as extremely important given the above studies regarding the theoretical properties of stochastic matrices and given the type of data one can expect [7] to utilize. Our ideas are embodied in two papers [3,4] which essentially tell the same story, but the first for those concerned with soils, and the second for a general audience that may be interested in the method for other applications. A practical technique is developed for the derivation of transition probability matrices, given data, and the derived matrix is shown to computationally reproduce the data. We are likely to use these ideas to transform an in–house deterministic analysis code into one that can account for uncertain soil media. A further refinement of this experimental–based derivation of Markov transition matrices is suggested for more complicated behavior. In addition, use of such matrices as predictive tools would add insight regarding their robustness.

# THE MARKOV FRAMEWORK for MODELING UNCERTAINTIES in SOIL CONSTITUTIVE/DYNAMIC BEHAVIOR

## CONTENTS

## TECHNICAL ABSTRACT

The focus of this research effort is the construction of a mathematical framework to help us understand the role of uncertainties in the constitutive/dynamic behavior of soils. The approach chosen is phenomenological and probabilistic. The phenomenological/probabilistic model is a logical choice given the nature of the soil medium and the respective necessity for significant data. Soil behavior falls into the category of irreversible systems and is modeled as such. While the modeling of soil behavior is our goal, we present here some ideas on the necessary mathematical framework and attendant notes on application by which this is to be accomplished. It is recognized that there are many issues to be addressed and resolved before the goal is achieved, and that the first of these, a mathematical framework, will have only been touched by this work.

Markov theory is used to provide a probabilistic framework for the modeling of constitutive uncertainties in dynamic soil behavior. This theory is attractive because it is the probabilistic analog to the classical physics approach to dynamics, it is easily recast into a computational form and can be rationally linked to data, and because it coincides with one concept of a stochastic constitutive model. Only the homogeneous or stationary Markov model is considered here. It is expected that this assumption will not severely limit the applicability of the theory.

In addition to the Markov probabilistic framework, a statistical framework is needed to connect the theoretical model and the data. This connection falls within estimation theory. While possible approaches were maximum likelihood estimates, Bayesian inference techniques, and Maximum Entropy ideas, a graphical, data-based technique was adopted as providing the most information with a minimum of assumptions. This connection with data is based on the definition of the state space within which the transition matrix of the Markov model operates. The mutually exclusive and complete definition of a state space is, in general, a very difficult task, given the complexity of the physical problem being addressed, but the approach chosen here appears reasonable for processes where data is obtainable. Thus, our work may be roughly divided into a theoretical component [2,15,16] and an experiment-based component [3,4]. That is, (*i*) given a stochastic matrix, draw conclusions regarding the system it represents, and (*ii*) given specific system information in the form of experimental data, translate the data into the computational transition probability matrix.

Some preliminary technical aspects are summarized next: Our work deals with stochastic matrices of order $n$, where these are $n$x$n$ matrices having each element greater than or equal to zero and such that row sums equal 1; in symbols:

$$\mathbf{A} = [a_{ij}]_{i,j=1}^{n},\ a_{ij} \geq 0,\ \sum_{j=1}^{n} a_{ij} = 1.$$

Such matrices arise naturally in the study of Markov chains [9,10].

Our theoretical approach is mainly geometric rather than algebraic. Particularly, we conceive of a stochastic matrix of order $n$ in two different ways. Such a matrix $\mathbf{A}$ can be regarded as an operator on $\mathbf{R}^n$, i.e., if $v$ is a row vector in $\mathbf{R}^n$, then $v \rightarrow v\mathbf{A}$ defines a linear operator map $\mathbf{R}^n \rightarrow \mathbf{R}^n$. However, by virtue of the fact that $\mathbf{A}$ is stochastic, even more can be said. For, if we allow $\Delta$ be the simplex in $\mathbf{R}^n$ with vertices $e_i (i = 1, \ldots, n)$, where $e_i$ denotes a vector whose $i^{th}$ component is 1 and all others are zero, then $\mathbf{A}$ maps $\Delta$ into $\Delta$. A second way of regarding the stochastic matrix $\mathbf{A}$ is to suppose it to be a point in $\mathbf{R}^{n^2}$; we identify the $n^2$ elements of $\mathbf{A}$ with the $n^2$ components of a point in $\mathbf{R}^{n^2}$. Then all the stochastic matrices form a closed bounded

polytope in $\mathbf{R}^{n^2}$, which we denote here by $\Sigma$, a polytope being in dimension $n$ what polygons and polyhedrons are in dimensions 2 and 3, respectively. The properties of $\mathbf{A}$, which lead naturally to the classification of all stochastic matrices, are intimately related to the geometrical description of $\Sigma$.

**The foremost property of a stochastic matrix is its convergence characteristics.** Convergence can be defined with respect to the Euclidian norm in $\mathbf{R}^n$ or $\mathbf{R}^{n^2}$. In the first case, let $v_0$ be an arbitrary vector in $\Delta \subset \mathbf{R}^n$. Then we can study the existence of $\lim_{m\to\infty} v_0\mathbf{A}^m$. Alternatively, regarding $\mathbf{A}$ as a point of $\mathbf{R}^{n^2}$, we can examine the existence of $\lim_{m\to\infty} \mathbf{A}^m$. It has been demonstrated that both viewpoints are closely related. An important question addressed is the relationship between the convergence properties and the eigenvalues of a stochastic matrix $\mathbf{A}$. The main instrument for this is the Frobenius–Perron Theorem [18], for which a proof is given using geometrical methods. The main result is that, given an arbitrary vector $v_0$ in $\Delta$, $\lim_{m\to\infty} v_0\mathbf{A}^m$ exists if and only if $\mathbf{A}$ has, for eigenvalues, no roots of unity other than 1. However, to determine more precisely the convergence properties of a stochastic matrix, we examine the existence of, and characterize the properties of, $\lim_{m\to\infty} \mathbf{A}^m$. If $\lambda$ is an eigenvalue of $\mathbf{A}$, then $\lambda^m$ will be likewise of $\mathbf{A}^m$; thus, we find that $\lim_{m\to\infty} \mathbf{A}^m$, if it exists, will have for eigenvalues only 0 and 1. It will be therefore possible to classify limit stochastic matrices (matrices of the form $\lim_{m\to\infty} \mathbf{A}^m$) via the multiplicity of the eigenvalue 0 or 1. It may seem that this problem is complicated by the fact that a stochastic matrix need not be diagonalizable. (We may recall that a diagonalizable matrix $\mathbf{A}$ is defined by the existence of a non–singular matrix $\mathbf{V}$ and a diagonal matrix $\Lambda$ such that $\mathbf{VAV}^{-1} = \Lambda$.) But, we show that the eigenvalue 1 always occurs in a diagonalizable manner, while 0 occurs likewise in limit stochastic matrices.

Also of interest is the class of nearly–completely decomposable stochastic matrices. Briefly, a matrix is said to be completely decomposable if, by renumbering if necessary, it can be written in the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & & & \\ & \mathbf{A}_2 & & \\ & & \ddots & \\ & & & \mathbf{A}_k \end{bmatrix},$$

where the $\mathbf{A}_i$ are matrices which run down the diagonal of $\mathbf{A}$, all other elements being 0. Of course, if $\mathbf{A}$ is stochastic, each $\mathbf{A}_i$ will be likewise; hence, a study of $\mathbf{A}$ simply reduces to the study of smaller matrices, $\mathbf{A}_i, (i = 1, \ldots, m)$. Courtois [8] has studied stochastic matrices that have their elements close to completely decomposable matrices. Apparently, these "nearly–completely decomposable" matrices have many modeling applications using Markov processes. In connection with such matrices, we prove the following result: If $\mathbf{B}$ is an arbitrary limit matrix, $\mathbf{B} = \lim_{m\to\infty} \mathbf{A}^m$, where $\mathbf{A}$ is a stochastic matrix, then there exists a nearly–completely decomposable stochastic matrix $\mathbf{B}_1$, such that $\mathbf{B} = \lim_{m\to\infty} \mathbf{B}_1^m$, and in fact $\mathbf{B}_1$ may be chosen arbitrarily close to a completely decomposable matrix.

Complementing the above work on the properties of stochastic matrices, an experimental–based procedure is developed to derive Markov transition probability matrices. For a given experiment data base, a state space is defined and used to derive transition probabilities by applying the frequency interpretation of probability. This approach can, in principle, be used regardless of the complexity of the behavior, to whatever accuracy necessary using a more refined state space definition. The resulting transition probability matrix is representative of the behavior of the medium, that is, the data from which it was derived. It is expected that the matrix is reasonably

robust and representative for "similar" types of soils.

## PROBLEM STATEMENT

The focus of this research effort is the construction of a mathematical framework to help us understand the role of uncertainties in the constitutive/dynamic behavior of soils. (For some background on the constitutive modeling of soils, see [1].)

A constitutive model or equation describes the macroscopic behavior resulting from the internal constitution of a material; basically, the model characterizes the individual material and its reaction to applied loads. A stochastic constitutive equation therefore characterizes the material where uncertainties exist about its material and geometric properties, and its reaction to applied loads. As a simple case, consider an ideal elastic solid undergoing axial strain resulting in axial stress according to the constitutive relation $\sigma = E\epsilon$ where $E$ is Young's modulus of elasticity. If $E$ is a random variable (or a field, that is, a function of position), then one or both $\sigma$ and $\epsilon$ can at best be described using probabilistic descriptors. If $\sigma = E\epsilon$ is incorporated into an equation of motion, then a stochastic constitutive dynamic model is created by way of the random properties of $E$.

One may view a generalized stochastic constitutive equation as one which probabilistically defines how a material undergoes dynamic behavior. A constitutive model must accurately describe the experimental data used to specify it, as well as predict behavior under conditions not covered by the original data. In the development of a constitutive model, it must be recognized that experimental data *is always difficult and sometimes impossible* to obtain. Therefore, the model should contain as few parameters requiring evaluation as possible. Furthermore, the model should be a function of the major system variables so that the underlying physics of the media/process is adequately represented, and thus hopefully understood.

Of course, no single constitutive model can represent any material in all situations. Even water, which is probably the most studied and best understood real material known to man, is never described by a single constitutive law to cover all situations. Thus, whenever a constitutive model is developed, only those features of material behavior relevant to the physics at hand should be included. Any completely general formulation, while philosophically pleasing, will only be of a formal nature.

Our effort has centered about the development of phenomenological, as opposed to microstructural, stochastic constitutive models for soils. Phenomenological models require data not only to determine parameter values, but also to validate the model itself. Such model building uses the following procedure [5]:

1. a generic form of the model is chosen,
2. one examines families that possess the general features known about the phenomena, and picks the class of models that encompass the phenomena in a reasonably complete manner,
3. evaluate the parameters of the model,
4. since the phenomenological model is probabilistic, results will also be probabilistic: probabilities of specific events, averages, sample function behavior,
5. validate the model using a variety of data.

The phenomenological model provides an ideal framework where concern exists about the correlation of real-life data with the predictive accuracy of the theoretical model. It also permits

the analyst to retain the theoretical model while adopting the new information that becomes available with more recent data.

## AN APPROACH to a COHERENT FORMULATION

Markov theory has been chosen to provide the framework for modeling constitutive uncertainties in soil behavior. There are several basic reasons for the adoption of this powerful theory in these studies:

1. The fundamental axiom upon which all Markov theory rests is the following: the probability that a system occupies some state at time $t_1$ later than time $t$ depends on its disposition at time $t$ and not on any time earlier than $t$. This is analogous to the approach of classical (deterministic) physics where the evolution of a system is established given its present state. Governing differential equations are such an example.
2. The need to know only the most recent state of the system is attractive also from the computational point of view. It would be possible to update the state of the system while retaining data from only the previous state. The complete history is not needed since it is effectively incorporated in this most recent information. This property is also valuable in translating experimental data into model parameters, as is discussed below.
3. Finally, the Markov state transition matrix appears to be an ideal mathematical counterpart to a concept of a stochastic constitutive model, where this transition matrix and corresponding probability states can be interpreted as a stochastic constitutive model coupled with the evolutionary system dynamics.

The Markov framework is a conceptual structure that we use to keep an, albeit sophisticated, accounting of how uncertainties in the system or inputs propagate in space and time. As a theoretical construct it is well understood. When it is used to gain physical understanding about a specific system, such as a soil, it is necessary that certain parameters particular to that system be incorporated into its mathematical structure. These parameters will be representative of the stochastic material and geometric characteristics of the system being studied. In summary, the proposed model is of two parts:

1. a probabilistic theoretical framework, and
2. a data–based parameter estimation technique.

Item 1 reflects our understanding of how a soil medium responds to its environment; we adopt Markov theory. Item 2 reflects the need to connect the theory to specific problems. This means the utilization of data.

It is noted that the behavior of a soil medium, under realistic conditions, is irreversible. This means that it is not possible to regain earlier states except in very limited cases. This understanding has an effect on the mathematical modeling, and the interpretations of predictions.

## THE GEOMETRIC THEORY of STOCHASTIC MATRICES

### Stochastic Matrices as Operators on a Simplex

Here, the basic properties of stochastic matrices conceived as linear operators acting on $\mathbf{R}^n$ are defined and proved. Actually, it is seen that the stochastic matrices do even better: they operate on a convex compact simplex of dimension $n-1$, which is contained in $\mathbf{R}^n$. In fact, this simplex is

the set of stochastic vectors, i.e., the set of probability distributions of a system having $n$ possible states.

Two methods are involved in this study. The first is to identify the simplest types of stochastic matrices, determine their properties, and then deduce the properties of the general stochastic matrices as built up from these simplest units. (The simplest stochastic matrices are the indecomposable ones.) The second method is to relate the properties of stochastic matrices to their eigenvalues.

The definition of decomposability can be motivated as follows. The numbering of the components of a vector is completely arbitrary; hence, the significant properties of a matrix which represents an operator on vectors should be defined so as to be independent of such numbering. Specifically, let $v$ be a row vector in $\mathbf{R}^n$ and suppose $v'$ is obtained from $v$ by some permutation of the components. We can write this as $v = v'\mathbf{T}$, where $\mathbf{T}$ is an $nxn$ matrix having only 0 and 1 as its elements; in fact, $\mathbf{T}$ is doubly stochastic and has an inverse, which is also a stochastic matrix. If $\mathbf{A}$ is a general stochastic matrix and $w = v\mathbf{A}$, we can write this equation in terms of permuted vectors $w', v' : w' = v'\mathbf{TAT}^{-1}$. Thus, we see that any significant property of $\mathbf{A}$ should also be shared by $\mathbf{TAT}^{-1}$, where $\mathbf{T}$ runs through all possible permutation matrices; note that these matrices $\mathbf{TAT}^{-1}$ will also be stochastic.

The definition of indecomposability follows from the above criterion: a stochastic matrix $\mathbf{A}$ is decomposable if there exists some permutation matrix $\mathbf{T}$ such that $\mathbf{TAT}^{-1}$ has the form

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ 0 & \mathbf{A}_{22} \end{bmatrix},$$

where $\mathbf{A}_{11}, \mathbf{A}_{22}$ are $n_1$ and $n_2$ order square matrices, such that $n_1 > 0, n_2 > 0$. $\mathbf{A}_{22}$ must necessarily be stochastic, but $\mathbf{A}_{11}$ is so if and only if $\mathbf{A}_{12} = 0$. By means of such permutation matrices, it is possible to assume $\mathbf{TAT}^{-1}$ to be of the form

$$\begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ 0 & & \mathbf{A}_r \end{bmatrix},$$

where $\mathbf{A}_{k+1}, \ldots, \mathbf{A}_r$ are stochastic and indecomposable. (We begin with the subscript $k+1$ since there may be $k$ decomposable and stochastic matrices.) Indecomposable, stochastic matrices are the easiest to study since they turn out to have a simple eigenvalue equal to 1. If no other eigenvalue has absolute value 1, then the matrix has very simple asymptotic properties which are investigated. More generally, the properties of an arbitrary stochastic matrix can be inferred from its indecomposable components.

The above definition of indecomposability was an algebraic one. There is a geometric equivalent. It was noted before that a stochastic matrix mapped a simplex of $\mathbf{R}^n$ into itself; we can define a stochastic matrix by this property. Further, we show that the property of indecomposability translates into the property that no proper part of the boundary of this simplex is mapped into itself. This alternative definition will often turn out to be more useful.

Four theorems on the eigenvalue properties of stochastic matrices have been stated and proved [15].

### Representation of Stochastic Matrices as Convex Polytopes

In the last section, stochastic matrices were viewed as operators acting on vectors, more particularly acting on stochastic vectors. However, it is desirable, when considering the classification of stochastic matrices, to view them also as vectors with $n^2$ components. In fact, it is necessary to

view them in this way to give meaning to statements such as: two stochastic matrices are close to one another. If these matrices are viewed as points in $\mathbf{R}^{n^2}$, the Euclidian distance or any topological equivalent enables us to determine the distance between stochastic matrices and to define continuous functions on the set of stochastic matrices. This then is the viewpoint taken in the following text.

It is then seen that the stochastic matrices form a simple figure in $\mathbf{R}^{n^2}$, namely, a convex bounded polytope. Also, those stochastic matrices which are limits of the form $lim_{m\to\infty}\mathbf{A}^m$ can be determined insofar as their geometrical status within this polytope. As a final step, the relation of nearly decomposable matrices to general stochastic matrices is examined. Eight theorems on stochastic matrices are stated and proved.

Consider, for example, the propagation of a single wave front traveling in one dimension, the real line, which is discretized into intervals of equal length. Each interval is labeled by an integer such that consecutive intervals bear consecutive labels. Let $v_i$ denote the probability that the wave front lies in the $i^{th}$ interval and is traveling to the right; $v_i'$ the probability it lies in the $i^{th}$ interval and is traveling to the left. Let $v_\ell$ be the row vector $(\ldots, v_{i-1}, v_i, v_{i+1}, \ldots)$ and $v_r$ similarly is given by $(\ldots, v_{i-1}', v_i', v_{i+1}', \ldots)$. Let $v = (v_\ell, v_r)$. Note that $\Sigma_i v_i + \Sigma_i v_i' = 1$, where all $v_i, v_i'$ are non-negative. Given $v$ at time $t_0$, we wish to calculate the transition matrix $\mathbf{R}$ such that $v\mathbf{R}$ will be the probability distribution of the wave front at the succeeding time interval $t_1$. Initially, we assume the velocity of propagation to right or left to be such that the wave front can travel the length of one interval in time $t_1 - t_0$.

Further, we define that at each boundary between intervals the wave front has probability $\rho$ of being reflected and $\tau$ of being transmitted; so that $\rho, \tau$ are non-negative and add to 1. Suppose $e_i$ denotes the row vector whose components are all zero except the $i^{th}$ which equals 1. If $v = (e_i, 0)$ at $t = t_0$, then $v = (\tau e_{i+1}, \rho e_i)$ at $t = t_1$. In a like manner, it is easy to see that $(0, e_i)$ is transformed to $(\rho e_i, \tau e_{i-1})$. Thus, we see that $\mathbf{R}$ is given by:

$$\mathbf{R} = \begin{bmatrix} \tau\mathbf{T} & \rho\mathbf{I} \\ \rho\mathbf{I} & \tau\mathbf{T}^{-1} \end{bmatrix},$$

where $\mathbf{I}, \mathbf{T}$ are matrices indexed by $\mathbf{Z}\mathbf{x}\mathbf{Z} : \mathbf{I}_{jk} = \delta_{jk}$, the Kronecker $\delta$, $\mathbf{T}_{j,k} = \delta_{j+1,k}$, $\mathbf{T}^{-1}_{j,k} = \delta_{j-1,k}$. Note that $\mathbf{T} \cdot \mathbf{T}^{-1} = \mathbf{T}^{-1} \cdot \mathbf{T} = \mathbf{I}$, so that the notation follows standard convention.

If we now introduce the probabilities $\{p_k\}$, $\sum_{k=1}^{\infty} p_k = 1$, where $p_k$ is the probability that the velocity of propagation is $k$ intervals per unit time, we obtain the transition matrix

$$\mathbf{Q} = \sum_{k=1}^{\infty} p_k \mathbf{R}^k.$$

A number of practical comments are due here:

1. the probability distribution of propagation velocity will be a function of the uncertainties associated with material properties, which, in turn, will depend on the type and magnitude of loading; as the material undergoes irreversible deformations, the distribution generally changes,
2. reflection and transmission coefficients $\rho$ and $\tau$ are measures of material homogeneity and geologic structure, and
3. we expect that for larger $k$ (extreme velocities with propagation over many elements), $p_k$ will become exceedingly small and practically negligible when compared to the early terms in the series; thus, the infinite series can be effectively truncated, and in some applications, after just a relatively few terms.

We resume the above development by calculating the powers of $\mathbf{R}$ as follows: Let

$$\mathbf{U} = \begin{bmatrix} \mathbf{T} & 0 \\ 0 & \mathbf{T}^{-1} \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 0 \end{bmatrix}, \quad \bar{\mathbf{I}} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{I} \end{bmatrix};$$

then, clearly $\mathbf{W}^2 = \bar{\mathbf{I}}$, and

$$\mathbf{WU} = \mathbf{U}^{-1}\mathbf{W},$$

where

$$\mathbf{U}^{-1} = \begin{bmatrix} \mathbf{T}^{-1} & 0 \\ 0 & \mathbf{T} \end{bmatrix}.$$

Thus,

$$\begin{aligned} \mathbf{R} &= \tau\mathbf{U} + \rho\mathbf{W}, \\ \mathbf{R}^2 &= (\tau\mathbf{U} + \rho\mathbf{W})^2 \\ &= \tau^2\mathbf{U}^2 + \rho^2\bar{\mathbf{I}} + \tau\rho(\mathbf{U} + \mathbf{U}^{-1})\mathbf{W}, \\ \mathbf{R}^3 &= \tau^3\mathbf{U}^3 + \tau\rho^2(2\mathbf{U} + \mathbf{U}^{-1}) + [\tau^2\rho(\mathbf{U}^2 + \mathbf{U}^{-2} + \bar{\mathbf{I}}) + \rho^3]\mathbf{W}, \\ &\vdots \end{aligned}$$

etc.

It is possible to modify the above problem to obtain a finite order transition matrix. Consider now that the boundaries between the $-1$ and 0 intervals and the $n$ and $n+1$ intervals are impermiable; these boundaries permit reflection only. The transition matrix $\mathbf{R}_n$ for the intervals 0 to $n$ is now a stochastic matrix of order $2n$ of the form:

$$\mathbf{R}_n = \left[\begin{array}{cccc|cccc} 0 & \tau & & & \rho & & & \\ & 0 & \tau & 0 & & \rho & & 0 \\ & & \ddots & & & & \ddots & \\ 0 & & & \tau & & 0 & & \\ & & & 0 & & & & 1 \\ \hline 1 & & & & 0 & & & \\ & \rho & & 0 & \tau & 0 & & 0 \\ & & \ddots & & & & \ddots & \\ 0 & & & & & 0 & & \\ & & & \rho & & & \tau & 0 \end{array}\right].$$

Note that $\mathbf{R}_n$, and therefore also $\mathbf{Q}_n = \sum_{k=1}^{\infty} p_k \mathbf{R}_n^k$, where all $p_k$ are non-negative, and $\sum_{k=1}^{\infty} p_k = 1$, are doubly stochastic. Thus, barring exceptional cases,

$$\lim_{k\to\infty} v_0 \mathbf{Q}_n^k = \frac{1}{2n}(1, 1, \ldots, 1),$$

where $v_0$ is an arbitrary $2n$-stochastic vector, i.e., an arbitrary probability distribution of wave fronts. This means that after a sufficient interval of time, the wave fronts are evenly (uniformly) distributed in the medium, regardless of the initial distribution.

Two exceptional cases are as follows:

1. $\tau = 0,\ \rho = 1,$
2. $\tau = 1,\ \rho = 0.$

These cases can result in $\mathbf{R}_n$ being a permutation matrix. In case 1, $\mathbf{R}_n$ has $n$ linearly independent eigenvectors for eigenvalue 1, namely $(e_i, e_i)$ for $i = 1, 2, \ldots, n$. This means that the wave fronts in the $i^{th}$ interval at $t = 0$ is trapped there for all future time. In case 2, $\mathbf{R}_n^{2n}$ is the identity matrix; after $2n$ time steps the system resumes its initial distribution and is recurrent.

In the example, it is noted that the finite order transition matrices were doubly stochastic. A consequence of this was that regardless of the initial probability distribution, the limiting distribution was generally homogeneous. Of course, this result should come as no surprise since the transition matrix in each case was derived on the assumption of geometric homogeneity. Thus, **we should understand double stochasticity of the transition matrix to be a manifestation of geometrical homogeneity of the medium being modeled.** If this assumption is dropped, we would obtain a transition matrix which is not doubly stochastic.

Also, in the expression for $\mathbf{R}_n$, replace the $\tau's$ and $\rho's$ by $\tau_i$ and $\rho_i$, where $i$ varies, but such that $\rho_i + \tau_i = 1$; and so obtain a matrix which is not doubly stochastic. In fact, consider the extreme case where all motion to the left is restrained by perfectly reflecting barriers; thus, $\mathbf{R}_n$ has the form

$$\mathbf{R}_n = \left[\begin{array}{cccc|cccc} 0 & 1 & & & 0 & & & \\ & 0 & 1 & 0 & & 0 & & 0 \\ & & \ddots & & & & \ddots & \\ 0 & & & 1 & & 0 & & \\ & & & 0 & & & & 1 \\ \hline 1 & & & & 0 & & & \\ & 1 & & 0 & 0 & & & \\ & & \ddots & & & & \ddots & \\ 0 & & & & & & & \\ & & & 1 & & & & 0 \end{array}\right].$$

The unique stochastic eigenvector of $\mathbf{R}_n$ for eigenvalue 1 is $\frac{1}{2}(e_n, e_n)$; from this we conclude the expected result that as $t \to \infty$, all the wave fronts congregate in the farthest right subinterval. Note further that since $\mathbf{R}_n$ maps $(e_n, 0)$ to $(0, e_n)$ and vice versa, $\mathbf{R}_n$ is decomposable and has $-1$ for an eigenvalue so that $\mathbf{R}_n^m$ does not converge as $m \to \infty$.

## STOCHASTIC MATRICES and DYNAMIC PROBLEMS

Consider substochastic matrices of order $n$: $\mathbf{P} = (p_{ij})_{i,j=1}^n$, where all $p_{ij} \geq 0$ and $s_i = \sum_{j=1}^n p_{ij} \leq 1$. The last condition is that the row sums are less than or equal to 1. If all row sums equal 1, the matrix is said to be stochastic; if all row sums are less than 1, the matrix is said to be properly substochastic.

The row sums have a significant effect on the eigenvalues of $\mathbf{P}$, as can be seen from the following. Let $\lambda$ be the dominant eigenvalue of $\mathbf{P}$, the eigenvalue of maximum modulus. It is known that $\lambda \leq 1$; $\lambda = 1$ if $\mathbf{P}$ is stochastic. Suppose $x = (x_1, \ldots, x_n)$ is the left eigenvector belonging to $\lambda$, normalized so that $\sum_i x_i = 1$. Note that such a normalization is possible since, for the dominant eigenvalue, all $x_i \geq 0$. Then $x\mathbf{P} = \lambda x$, and taking row sums of both sides, $\sum_i x_i s_i = \lambda$. Therefore,

$$\mu = 1 - \lambda = \sum_{i=1}^n x_i(1 - s_i).$$

Only the terms for which $s_i < 1$ appear in the sum. If $\mathbf{P}$ is indecomposable, all $x_i > 0$ and we see again the fact that if $\mathbf{P}$ is indecomposable and properly substochastic, then the dominant eigenvalue $\lambda < 1$.

*Such matrices can arise from finite difference approximations of Laplace's equation.* (As such, applicability to a wide range of physical problems is seen.) Consider a square grid in two dimensions where the nodal points are denoted by $(i,j)$, where $i,j$ are positive integers. If $u$ is a function on this grid $\Delta u$ at the point $(i,j)$ is approximated by

$$\frac{1}{4h^2}[u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}],$$

where $h$ denotes the grid size. Setting $h = 1$, one sees that $-\Delta_c = \mathbf{I} - \mathbf{A}$, where $\mathbf{I}$ is the identity matrix and $\mathbf{A}$ is a stochastic matrix whose non-zero entries are usually $1/4$; the order $n$ of both matrices is the number of grid points. Suppose $\mathbf{A}$ is modified so that each boundary grid point is an absorbing point. Then $\mathbf{A}$ takes the following form:

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_d & 0 \\ \mathbf{R} & \mathbf{P} \end{bmatrix}.$$

Here $\mathbf{I}_d$ is the identity matrix of order $d,d$ being the number of boundary grid points, and $\mathbf{P}$ is a substochastic matrix of order $n - d$. In fact, $\mathbf{P}$ must be properly substochastic; otherwise, the boundary would not be connected to our region. The number of rows $d_1$ of $\mathbf{P}$ which have sum $< 1$ can now be interpreted as the number of grid points directly adjacent to the boundary. We should expect $d_1 \leq d$ with near equality. Ideally, equality should be an indication of smoothness of the boundary, for example, the absence of corners.

It will be noticed that other algebraic properties of $\mathbf{P}$ correspond to geometric properties of the region which is modeled by the grid. For example, failure of the region to be connected corresponds to a matrix $\mathbf{P}$ which is completely decomposable, that is, by renumbering indices we obtain a matrix of the form,

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & 0 \\ 0 & \mathbf{P}_{22} \end{bmatrix}.$$

Indecomposable matrices, those matrices which can be put in the form

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ 0 & \mathbf{P}_{22} \end{bmatrix},$$

where $\mathbf{P}_{12} \neq 0$, simply do not occur in this context. Additional properties of reversible matrices, and matrix reductions have been derived [16].

We can apply, for example, the foregoing to the solution of the following dynamical problem: let us consider one of the simplest examples in mechanics, the harmonic oscillator. Let $x(t)$ be a function for $t \geq 0$ which is a solution of the differential equation

$$\ddot{x} + \omega^2 x = 0,$$

where $x(0) = x_0$, $\dot{x}(0) = 0$, so that $x(t) = x_0 \cos \omega t$, $\dot{x}(t) = -\omega x_0 \sin \omega t$. If $\omega$ is held constant, $(x, -\dot{x})$ describes an ellipse in phase space.

Since the orbit is closed, we can approximate the system by a finite transition probability matrix as follows. Divide the curve into $n$ divisions and let $a_{ij}$ denote the probability the particle lands in the $j^{th}$ interval in a unit time supposing that it was in the $i^{th}$ interval just before the transition. If the intervals are numbered consecutively $1, 2, 3, \ldots, n$, so that after $n$ we regain 1,

then the matrix $\mathbf{A} = (a_{ij})_{i,j=1}^{n}$ should have the property that the matrix be the same regardless of which interval is numbered 1. Thus, we are led to conclude that $a_{i+1,j} = a_{i,j-1}$. Here the convention is adopted that the subscripts are taken mod $n$; e.g., $a_{21} = a_{1n}$. Thus, if the first row is given by $a_{1i} = p_{i-1}$, for $i = 1, \ldots, n$, all other rows are given by $a_{ij} = p_{j-i}$ (again with the mod $n$ convention for subscripts). If $\mathbf{T}$ is the $n^{th}$ order matrix defined by

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 & 0 & & & \\ 0 & 0 & 1 & 0 & \cdots & & \\ 0 & 0 & 0 & 1 & & & \\ & & & & \ddots & & \\ & & & & & & \\ 0 & 0 & & & \cdots & & 1 \\ 1 & 0 & & & & 0 & 0 \end{bmatrix},$$

then $\mathbf{A}$ is given by $\mathbf{A} = \sum_{i=0}^{n-1} p_i \mathbf{T}^i$. The eigenvalues of $\mathbf{T}$ are seen to be $\lambda_j = \exp(2\pi ij/n)$, where $j = 0, 1, \ldots, n-1$ and $i = \sqrt{-1}$. Indeed, if $x = (x_0, x_1, \ldots, x_{n-1})^T$ is a right eigenvector for $\mathbf{T}$, then $x_k = \exp(2\pi ijk/n)$. These must still be the eigenvectors for $\mathbf{A}$, but now the eigenvalues of $\mathbf{A}$ are

$$\lambda_j = \sum_{k=0}^{n-1} p_k \exp(2\pi ijk/n),$$

which displays each $\lambda_j$ as a mean of $n^{th}$ roots of unity. From this it is clear that two cases arise:

1. All $p_k$ except one are equal to 0. In this case $\mathbf{A}$ is periodic, i.e., $\mathbf{A}^m = \mathbf{I}$ for some $m$ which is a divisor of $n$. Thus, if $v_0$ is an initial probability distribution, it will recur after $m$ time steps: $v_0 = v_0\mathbf{A}^m$.
2. At least two $p_k > 0$. In this case $lim_{m\to\infty}\mathbf{A}^m$ exists, and in fact we must have $lim_{m\to\infty} v_0\mathbf{A}^m = n^{-1}(1, 1, \ldots, 1)$ for any arbitrary initial probability distribution $v_0$. In other words, as time goes to infinity we must approach a state where there is equal probability of the particle being in any of the $n$ subintervals.

Note that case 1 is in fact the deterministic case. In the above formulation the only relevant numerical data is the probability distribution $(p_0, \ldots, p_{n-1})$ which can be directly calculated from the probability distribution for the natural frequency (or spring constant) of the system. This can be said to govern the speed at which the particle goes around the orbit: $\mathbf{T}^k$ is the matrix which corresponds to a particle moving through $k$ subintervals in a unit time. Nothing is said regarding the radius of the orbit (actually one should more strictly speak of the semi-major axis since the orbit is an ellipse). The radius of the orbit is a function of the energy, which is constant if no external force is applied. Suppose now there is such a force applied. Then we may suppose its effect is to alter the radius of the orbit. Since the loading probability may be assumed independent of the structural properties of the system, we may solve this problem independently of the former problem; i.e., the radial distribution is independent of the tangential distribution.

An example of a stochastic matrix of order $n + 2$ for such a transition of orbit in unit time is given by

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & \cdots & & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & \cdots & & & & \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \cdots & & & \\ & & & & \cdots & & & \\ & & & \cdots & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & \cdots & & 0 & 0 & 1 \end{bmatrix}.$$

The meaning of **Q** is as follows: there exist $n$ adjacent orbits and two absorption states such that a particle has equal probability of transferring to one of its two adjacent orbits or to an absorption level, from which there is no return. Then $\lambda = 1$ is an eigenvalue of **Q** with multiplicity 2; corresponding left eigenvectors can be taken as $(1,0,\ldots,0)$ and $(0,\ldots,0,1)$. The remaining eigenvalues of **Q** are the eigenvalues for the portion of **Q** which is marked off in the dashed box. Applying to **P**, the same method employed earlier, we see that the principal right eigenvector for **P** must be of the form, $a = (a_1,\ldots,a_n)^T$, where $a_k = \sin(k\pi/(n+1))$, where the corresponding eigenvalue is given by $\lambda = \cos(\pi/(n+1))$. In fact, the totality of eigenvalues and right eigenvectors for **P** is given by replacing $\pi/(n+1)$ by $j\pi/(n+1)$, for $j = 1,2,\ldots,n$. Thus if $\lambda_1$ is the dominant eigenvalue, $-\lambda_1$ is also an eigenvalue, and actually, all eigenvalues occur with their negative values. This could also have been seen from the fact that the trace of **P** equals 0. Because of this, it is not possible to use our previous theory to associate a probabilistic meaning to the eigenvector associated to $\lambda_1$.

But this shortcoming can be remedied very easily by considering $\mathbf{P}^2$ instead of **P**. $\mathbf{P}^2$ has positive eigenvalues but, except for the possible $\lambda = 0$, all occur with multiplicity 2. However, **P** is completely decomposable and it suffices to work with only one of the "parts" of **P**. The physical interpretation illuminates the foregoing algebraic assertion: in two time steps, the system must lie in an even numbered state, if it started out in an even numbered state, with the same if we replace "even" by "odd". Thus, if we double our time step and consider only the even numbered states we obtain a transition matrix, which we again call **P**:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 0 & & \cdots & & \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & \cdots & & \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & \cdots & \\ & & & \cdots & & & \\ & & & & \cdots & 0 & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

This new **P** should have an order approximately half that of the previous **P**, but again we may suppose the order to be $n$. The dominant eigenvalue is now given by

$$\lambda = \frac{1 + \cos(\frac{\pi}{n+1})}{2} = \cos^2 \frac{\pi}{2(n+1)}.$$

The corresponding left (or right) eigenvector is given by the components $a_k = \sin(k\pi/[n+1]), k = 1,2,\ldots,n$.

If the absorption states are removed and $n$ permitted to go to infinity, it will then be found that the limiting distribution approaches that of a Gaussian or normal distribution.

## MARKOV CHAIN TRANSITION MATRICES as CONSTITUTIVE MODELS of SOIL DYNAMIC BEHAVIOR

To apply a Markov chain model, we recall that the transition probabilities $\{X_n\}$ with state space $\{0,1,2,...\}$, best exhibited in the form of a matrix

$$P(m,n) = \begin{bmatrix} p_{0,0}(m,n) & p_{0,1}(m,n) & \cdots & p_{0,k}(m,n) & \cdots \\ p_{1,0}(m,n) & p_{1,1}(m,n) & \cdots & p_{1,k}(m,n) & \cdots \\ & & \cdots & & \cdots \\ p_{j,0}(m,n) & p_{j,1}(m,n) & \cdots & p_{j,k}(m,n) & \cdots \\ & & \cdots & & p_{J,K}(m,n) \end{bmatrix},$$

must be derivable. The elements of a transition probability matrix $P(m,n)$ satisfy the conditions

$$p_{j,k}(m,n) \geq 0 \quad \text{for all } j,k$$
$$\sum_k p_{j,k}(m,n) = 1 \quad \text{for all } j.$$

One may derive the recursive relation:

$$p(n) = p(0)P(0,n),$$

where

$$p(n) = [p_0(n)\ p_1(n) \ldots p_j(n) \ldots],$$
$$p_j(n) = P[X_n = j].$$

It follows that the probability law of a Markov chain $\{X_n\}$ is completely determined once one knows the transition probability matrices, and the unconditional probability vector $p(0)$ at time 0. In the case of a **homogeneous** (time–invariant) Markov chain $\{X_n\}$, let

$$P(n) = \{p_{j,k}(n)\}, \quad P = \{p_{j,k}\}$$

denote respectively the $n$–step and the one–step transition probability matrices. From the above equations, it is observed that

$$P(n) = P^n,$$
$$p(n) = p(0)P^n.$$

Consequently, the probability law of a homogeneous Markov chain is completely determined once the one–step transition probability matrix $P = \{p_{j,k}\}$ is known, and the unconditional probability vector $p(0) = \{p_j(0)\}$ at time 0.

Markov theory is a conceptual framework for modeling the propagation in time and space of system and environmental uncertainties. As a theoretical construct it is extensive [6,14,23]. When it is used to gain physical understanding about a specific dynamical system, it is necessary that certain parameters particular to that system be incorporated into its mathematical structure. These parameters will represent the stochastic material, geometric, and environmental characteristics of the system being studied. The transition probability matrix is the vehicle by which such information is incorporated into the Markov model.

Others have taken the Markov chain approach to physical modeling [5,19, for example]. However, the text below, to the best knowledge of this author, is not available elsewhere. An Appendix summarizes some of the important definitions from Markov theory.

### Definition of States

Implicit in the above discussion is the fact that the Markov model operates on a system which has been partitioned into states. Inherent in the definition of the state space for a problem is an understanding of, even if only in general terms, what are the likely system response patterns to the environments to which it will be exposed. The state space is mutually exclusive and complete; all possible states can represented in a unique fashion. It is desirable to devise a procedure for state–space specification that is unbiased, meaning that it does not impose the modeler's **a priori**

expectations on the types of behavior the model can predict. Ideally, one would aim for a model mechanism which would generate and augment the state space using initial and current state information.

The process of state space definition will begin with the determination of behavior regimes that are of particular importance or interest. Just as with a finite element model, where more elements are specified in regions of complexity or possible critical behavior, the definition of a state space exhibits the analyst's expectations. Thus, if one is defining the state space for a material that is to undergo significant strain deformation beyond the yield strain, $\varepsilon_y$, one might define a relatively small number of states in the elastic region and a larger number in the yielded region. Uncertainty regarding the numerical value of $\varepsilon_y$ will require the definition of several yield states with corresponding transition probabilities according to a probability distribution law derived from data. Similarly, uncertainty about the elastic loading modulus may motivate the definition of several loading moduli.

Theoretically, the state definition procedure is not a trivial task since the analyst must anticipate behavior states that may not have been previously encountered, and do this without hopelessly exceeding analytical and/or computational abilities. It is also not an isolated task, but one which must be carried out with knowledge about the ease of estimating the transition probabilities. It is only useful to define states which can be observed experimentally, and for which transition probabilities can be estimated.

In particular to the procedure proposed herein, a relatively straightforward definition of a state space and derivation of transition probabilities based on experimental data is demonstrated. The following example displays how test data can be used to help specify the state space for a process. Figure 1 depicts the results of a series of 10 dynamic uniaxial strain tests for undisturbed specimens of CARES–Dry sand [7]. Differences in response can primarily be attributed to (small) differences in dry density. It is noted that the stress–strain curves are geometrically similar, but spatially translated along the strain axis.

A possible state space could be obtained by the superposition of a square/rectangular grid on the above figure of experimental data. This approach is described below.

### Estimation of Transition Probabilities From Data

The main thrust of this section is now developed; that is, given experimental data about a certain class of soil behavior, what approach can be used to derive a transition probability matrix that is a reasonable measure of the evolutionary properties of this soil media class?

Consider the generic, parametric $xy$ curves of Figure 2. It is useful to be able to derive the transition probability matrix for the process which results in this behavior, primarily in the theory that such behavior is in some useful sense representative of near–media behavior. The frequency interpretation of probability will be used here. This approach assumes that, given multiple realizations of a process (as one would have due to a sequence of experiments), as the number of realizations becomes large, the ratio of the number of a specific realization to the total number of realizations is a reasonable measure of the probability of the specific occurrence. This approach assumes also that each realization is equally probable, although one could appropriately weigh more likely outcomes.

Thus, if one is interested in estimating the, here time–invariant, transition probability between state $i$ to state $j$, $\pi_{ij}$, one could use the relation

$$\pi_{ij} = \frac{\sum i \to j}{\sum all\ i},$$

where the denominator represents all transitions from $i$, including $i \rightarrow j$, and those that are absorbed within $i$. Using this procedure, for the generic experimental data of Figure 2, the following transition probability matrix $\Pi$ can be derived using a graphical procedure:

$$\Pi = \begin{array}{c} \\ 0 \\ 11 \\ 12 \\ 13 \\ 14 \\ 22 \\ 23 \\ 24 \\ 32 \\ 33 \\ 34 \\ 43 \\ 44 \end{array} \begin{array}{c} \begin{array}{cccccccccccccc} 0 & 11 & 12 & 13 & 14 & 22 & 23 & 24 & 32 & 33 & 34 & 43 & 44 \end{array} \\ \left( \begin{array}{cccccccccccccc} 3/3 & & & & & & & & & & & & \\ & 2/3 & & & & 1/3 & & & & & & & \\ & & 1/2 & & & 1/2 & & & & & & & \\ & & & 2/3 & & & 1/3 & & & & & & \\ & & & 1/2 & 1/2 & & & & & & & & \\ & & & & & & 1/2 & & 1/2 & & & & \\ & & & 1/3 & & & & 1/3 & & 1/3 & & & \\ & & & & 2/3 & & & & & & 1/3 & & \\ & & & & & & & & & 1/1 & & & \\ & & & & & & 1/3 & & & & 1/3 & 1/3 & \\ & & & & & & & 2/3 & & & & & 1/3 \\ & & & & & & & & & 1/1 & & & \\ & & & & & & & & & & 1/1 & & \end{array} \right) \end{array},$$

where the left column label represents the initiating state ("from"), and the top row label represents the receiving state ("to"). It is assumed that from the initial condition 0, all transitions are to 11. Thus, the $\frac{3}{3}$ entry under column 11.

As an example of the calculations, consider the transitions from state (2,2): there are 2 paths in (2,2), one goes to (3,2), the other to (2,3); thus, each is assigned a probability of $\frac{1}{2}$.

An implicit assumption that is made above is that the time increments between transitions are of such a duration that transitions will only occur between adjacent (bordering) states. This requirement can always be satisfied by appropriately refining the state space. In general, each state may lead to a transition to any of 8 states for the uniform state space grid of Figure 2.

The matrices considered here are sparse. For most experimental data, this will be the case. Note the banded nature of the non–zero elements. This is also a result of the manner in which the state space is numbered. It is emphasized that

1. the derivation of the above transition matrix does not take into account whether the individual curves are in a "loading" or "unloading" phase; this distinction will be important for some applications, and is discussed in the next section, and
2. while the transition probability matrix is *homogeneous*, an implicit time–scale exists due to the assumption that during one time increment only the adjacent states may be entered.

### Model Refinement for More Complicated Behavior

Here, it is of interest to model behavior such as the reversal of trends witnessed in stress–strain behavior of materials subjected to loading and unloading cycles. The procedure of the previous section is modified by augmenting the state space such that "loading" behavior is distinguished from "unloading" behavior. Thus, for example, in Figure 2 state (3,3) is in actuality 2 states: $(3,3l)$ and $(3,3u)$, where the suffix $l$ denotes loading, and $u$ denotes unloading.

With this approach, one can partition the transition probability matrix into 4 sub–matrices for the problem at hand:

$$\Pi = \begin{array}{c} \\ l \\ u \end{array} \begin{array}{c} \begin{array}{cc} l & u \end{array} \\ \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \end{array},$$

where **A** represents an evolution from loading to continued loading, **B** represents an evolution from loading to unloading, **C** represents an evolution from unloading to loading, and **D** represents an evolution from unloading to continued unloading.

For the generic curves of Figure 2, matrix $\Pi$ is 18 rows by 18 columns, and the sub–matrices are as follows:

$$\mathbf{A} = \begin{array}{c} \\ 0 \\ 11 \\ 12 \\ 13l \\ 22 \\ 23l \\ 24l \\ 32 \\ 33l \\ 34l \end{array} \begin{array}{c} \begin{array}{cccccccccc} 0 & 11 & 12 & 13l & 22 & 23l & 24l & 32 & 33l & 34l \end{array} \\ \begin{pmatrix} 3/3 & & & & & & & & & \\ & 2/3 & & & 1/3 & & & & & \\ & & 1/2 & 1/2 & & & & & & \\ & & & & & 1/1 & & & & \\ & & & & & 1/2 & & 1/2 & & \\ & & & & & & 1/2 & & 1/2 & \\ & & & & & & & & & 1/1 \\ & & & & & & & & 1/1 & \\ & & & & & & & & & 1/2 \\ & & & & & & & & & \end{pmatrix} \end{array},$$

$$\mathbf{B} = \begin{array}{c} \\ 0 \\ 11 \\ 12 \\ 13l \\ 22 \\ 23l \\ 24l \\ 32 \\ 33l \\ 34l \end{array} \begin{array}{c} \begin{array}{cccccccc} 13u & 14u & 23u & 24u & 33u & 34u & 43u & 44u \end{array} \\ \begin{pmatrix} & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & 1/2 & \\ & & & & & & & 1/1 \\ & & & & & & & \end{pmatrix} \end{array},$$

$$\mathbf{D} = \begin{array}{c} \\ 13u \\ 14u \\ 23u \\ 24u \\ 33u \\ 34u \\ 43u \\ 44u \end{array} \begin{array}{c} \begin{array}{cccccccc} 13u & 14u & 23u & 24u & 33u & 34u & 43u & 44u \end{array} \\ \begin{pmatrix} 2/2 & & & & & & & \\ 1/2 & 1/2 & & & & & & \\ 1/1 & & & & & & & \\ & 2/2 & & & & & & \\ & & 1/1 & & & & & \\ & & & 2/2 & & & & \\ & & & & 1/1 & & & \\ & & & & & 1/1 & & \end{pmatrix} \end{array},$$

and $\mathbf{C} = 0$ in this instance, where it is noted that these sub–matrices are sub–stochastic, that is, their row sums are $\leq 1$. Of course, the row sums of transition matrix $\Pi$ are equal to 1; $\Pi$ is a stochastic matrix. In the above, all rows/columns exclusively composed of $0's$, except for the "0" column, have been removed.

It is emphasized that the above transition probability matrix is only representative of the "data curves" of Figure 2, and is a product of the state space assumed in that figure. This means that one will be able to reproduce these curves, using the derived transition probability matrix, by applying the theory as outlined in the previous section. Thus, one would obtain the following state probabilities at each increment in the model:

$$
\begin{array}{lcccccccccccccccccc}
p(0) = & [1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0] \\
p(1) = & [0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0] \\
p(2) = & [0 & 0 & 0.67 & 0.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0] \\
p(3) = & [0 & 0 & 0 & 0.33 & 0.33 & 0.17 & 0 & 0.17 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0] \\
p(4) = & [0 & 0 & 0 & 0 & 0 & 0.50 & 0.08 & 0.17 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0] \\
\vdots & & & & & & & & \vdots & & & & & & & & & \\
p(11) = & [0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.72 & 0.28 & 0 & 0 & 0 & 0 & 0 & 0] \\
\vdots & & & & & & & & \vdots & & & & & & & & & \\
p(16) = & [0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.99 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0]
\end{array}
$$

Such state information would be useful in a comprehensive computational tool where the evolutionary behavior of soil samples is required. It is conceivable that tables of such information can be developed for different samples under different environmental conditions. As such, a look–up procedure can be established in the main program.

The hope, and this is the subject of further work, is that the given data is also representative of process behavior in some "near field", the definition of which will be dependent on the correlation structure of the media. To assess this property, it will be necessary to obtain similar data for field points near the original point, and to establish procedures for estimating such correlations. Given the results of the present work, and of the necessary extensions, it is believed that a framework for modeling the evolutionary behavior of a media such as soil is possible using this phenomenological point of view. In a general formulation, each state may be decomposed into "loading" and "unloading" states, **or whatever behavior class is most appropriate.**

This approach to estimating the transition probability matrix can be used in problems where not much data is available, but one can draw on engineering judgment to set bounds on the behavior and loading. In this instance, one may draw upper and lower bound $\sigma - \varepsilon$ curves and "spread" the probabilities evenly (or with some weighting) throughout the state space. Such a transition matrix may provide a rough measure of the evolutionary soil behavior.

The state space definition, a rectangular grid above, is purely a construct of the analyst, and cylindrical, spherical, or other geometries may be more useful for other types of problems. The transition matrix may, of course, be used to represent a 3–dimensional problem. The key is always an adequate data set, or other perhaps subjective information, for estimating probabilities.

Applying this graphical procedure to the experimental CARES–Dry sand data is straightforward, but, due to the 10 samples, one may superimpose a refined grid on the sample paths, which are overlapping and "close". Transition probabilities for this application will be in multiples of $\frac{1}{10}$. The state space grid need not be uniform, and may be more refined in regions of critical behavior. Again, the transition matrix will be banded and sparse. If these sample paths are characteristic of the underlying soil media, then the derived transition probability matrix will be representative of the soil behavior.

For behavior which is more complicated than that presented above, such as cyclical, hysteretic behavior of soils, one may extend the state space definition such that each complete loading–unloading cycle is represented by a state–space in two dimensions, and each subsequent loading–unloading cycle extends the state space in the third dimension. Thus, a three dimensional grid will be representative of soil undergoing many cycle hysteresis. In such a manner, so–called "memory–dependent" behavior can be accounted for in the model. The key to the success of such an approach is the availability of sufficient data. But then, this is the key to any model purporting to be representative of some aspect of the physical world.

In soil dynamics application, the analyst would use the Markov chain state transition model to estimate the "location" of the soil in its state space. This information, i.e., $p(0),\ldots,p(16)$ above, would be fed back to the dynamics computer code whenever such information is needed to update the system dynamics. Some thought would be needed as to whether the most probable state is used, or some average of possible states be used, to define the evolution of the soil in its state space. Use of a particular transition matrix presupposes that a similar soil, and loading, exist for the problem at hand. It is intended that the transition matrix derivation be automated; given experiment-generated data points, an algorithm can be used to generate the transition probabilities for specific problems in real time.

## CONCLUDING REMARKS & SUGGESTIONS for ADDITIONAL WORK

We have learned much during the course of our work here. Certain aspects of our work, had we been able to retrace our steps, would have received less emphasis, and other aspects more emphasis. But, of course, that is a decision that can only be made *a fortiori*. Where we have made progress: a generic framework for modeling complex media such as soils is identified and explored; an approach to extract probabilistic information from experimental data is proposed and demonstrated using what we believe is the most comprehensive soil data available. Where it would have been interesting to have reached a more advanced stage: more progress would have been desirable in translating the mathematical understanding of Markov transition matrices into a physical understanding of the constitutive/dynamic behavior of the soil.

A continuation of this work would require the following components. The first is the extension of some of the theoretical lines drawn in categorizing stochastic matrices, their convergence properties, and the linking of matrix classes with dynamic behavior. In addition, the derivation of transition matrices which represent more complicated behavior, such as hysteresis, would demonstrate the validity of the graphical methodology presented in general. As part of this, testing the robustness of this methodology would be an enlightening activity. Automating the procedure so that transition matrices can be generated as data is compiled in an experiment could prove very useful in linking mathematical model to data.

One issue of prime concern, and one which becomes paramount to those who must work with complicated media such as soil, is the question of the modeling/predictive limits of analytical methods. That is, as good as our mathematical models and experimental techniques become, there is an inherent uncertainty in the medium that is of such significance that the results of any analysis, while rigorous, may not provide information which will be useful in a practical (predictive) sense (except perhaps over extremely short time durations).

The problem alluded to above is a function of the "scales" of the physical problem under study. For example, a site of 1,000 cubic feet may be reasonably modeled, given general information regarding the site environment, and some minimal tests within that 10 foot cube. This is because any discontinuities will have a reasonable chance of being identified, and that other "continuous changes" will not likely be significant due to the size of the length scale, that is, the length scale is small relative to any discontinuities.

If one now considers a cube with sides of length 1,000 foot ($1\text{x}10^9$ cubic feet), a different class of problem arises. This is because discontinuities can be completely contained within the volume, and may be missed by tests. The significance of the site environment is less important. In addition, continuous changes in properties now will work over a larger length scale. If one adds the fact that all variations in the medium have random components that are easily on the order of the mean value, then it can be concluded that any known model of the medium will not likely be representative of even adjacent sites.

This problem of "scales" of physical behavior is encountered in other disciplines, such as turbulence in oceanic and atmospheric processes, where parameter variations over the scales of the problem are significant, and therefore, predictions are limited. In fact, small scale turbulence results in large scale response. Similar difficulties exist for the problem at hand.

How does one transcend this problem? These concluding comments will not be a source for such resolution, rather, a few brief observations. Is the answer with microstructural modeling? We are not qualified to address this technical approach. However, if a model is microstructural instead of phenomenological, then the above problem of scales will tend to become more intricate. For, even if at a microscale an accurate mathematical model is derived, this model must be able to predict gross behavior at the larger scales of interest in application. To do this, the generalized forces and pressures that a microstructural model predicts at microscale must "accumulate" in some sense to the physical forces and pressures measurable in the laboratory and field. To do this, the microstructural model must be valid at many orders of magnitude of behavior, from the microscopic to the aggregate. Otherwise, the microstructural model will be only a formal expression.

Considering the difficulties of models which attempt predictions at much smaller behavior ranges and scales, it will be interesting to discover how microstructural models are approaching the difficulties briefly discussed above.

## REFERENCES[1]/BIBLIOGRAPHY


[1]* Benaroya, H, Sandler, I, and DiMaggio, F, "Uncertainties in Soil Constitutive Modeling," Weidlinger Associates Technical Report No 1, prepared for AFOSR under Contract No F49620-85-C-0045DEF, Sept 1985

[2]* Benaroya, H, "The Markov Framework for Modeling Uncertainties in Soil Constitutive-Dynamic Behavior, Part I: Overview," Weidlinger Associates Technical Report No 2 under Air Force Office of Scientific Research Contract No F49620-85-C-0045DEF, April 1986

[3]* Benaroya, H, "Markov Chain Transition Matrices as Constitutive Models of Soil Dynamic Behavior;" research supported by the Air Force Office of Scientific Research, Contract No F49620-85-C-0045DEF; presented as "Markov Chains & Soil Constitutive Modeling" at the ASCE EMD/GTD/STD Joint Specialty Conference on Probabilistic Mechanics and Structural and Geotechnical Safety, VPI&SU, Blacksburg, 25-27 May 1988; submitted to ASCE **J Eng Mech**

[4]* Benaroya, H, "Markov Transition Probabilities from Experimental Data;" research supported by the Air Force Office of Scientific Research, Contract No F49620-85-C-0045DEF; presented as "Markov Methods & Dynamic Models" at the ASCE EMD Specialty Conference, VPI&SU, Blacksburg, 23-25 May 1988; accepted for publication by **Appl Math & Computation**

[5] Bogdanoff, J L, and Kozin, F, **Probabilistic Modeling of Cumulative Damage**, John Wiley & Sons, 1985

[6] Box, G P, and Tiao, G C, **Bayesian Inference in Statistical Analysis**, Addison-Wesley, 1973

[7] Cargile, J D, "Geotechnical Investigation for the Cares-Dry Site: Laboratory Test Results," US Army Engineering WES Technical Report, Feb 1984

[8] Courtois, P J, **Decomposability - Queueing and Computer System Applications**, Academic Press, 1977

[9] Chung, K L, **Markov Chains With Stationary Transition Probabilities**, Second Ed, Springer-Verlag, 1967

[10] Cox, D R, and Miller, H D, **The Theory of Stochastic Processes**, Chapman and Hall, 1965

[11] Dodziuk, J, "Difference Equations, Isoperimetric Inequality and Transience of Certain Random Walks", Transactions of the American Mathematical Society, Vol 284, No 2, Aug 1984

[12] Doyle, P G, Snell, J L, **Random Walks and Electric Networks**, The Carus Mathematical Monographs, No 22, The Mathematical Association of America, 1984

[13] Dugundji, J, **Topology**, p341 and 224, respectively, Allyn and Bacon, 1966

[14] Fisz, M, **Probability Theory and Mathematical Statistics**, 3rd ed, Krieger, 1980


[1]An asterisk "*" next to a reference signifies that it is a report or paper written under the contract for which this is a final report.

[15]* Goldstein, G, Benaroya, H, "The Markov Framework for Modeling Uncertainties in Soil Constitutive–Dynamic Behavior, Part II: The Geometrical Theory of Stochastic Matrices," Weidlinger Associates Technical Report No 3 under Air Force Office of Scientific Research Contract No F49620–85–C–0045DEF, October 1986

[16]* Goldstein, G, Benaroya, H, "Stochastic Matrices and Dynamic Problems;" research supported by the Air Force Office of Scientific Research, Contract No F49620–85–C–0045DEF, in preparation for submission

[17]* Goldstein, G, Benaroya, H, "Fractal Modeling of Soil Media;" research supported by the Air Force Office of Scientific Research, Contract No F49620–85–C–0045DEF, submitted for publication to **Appl Math Modelling**

[18] Householder, A S, **The Theory of Matrices in Numerical Analysis**, p48, Dover, 1964

[19] Hsu, C S, **Cell–to–Cell Mapping**, Springer–Verlag, 1987

[20] Kac, M, "Random Walk and the Theory of Brownian Motion", American Mathematical Monthly, vol 54, No 7, republished in Nelson Wax, Selected Papers on Noise and Stochastic Processes, Dover Publications, 1954

[21] Lang, S, **Algebra**, p398, Addison–Wesley, 1965

[22] Minc, H, "Permanents, Encyclopedia of Mathematics and its Applications," vol 6, Addison Wesley, 1978

[23] Parzen, E, **Stochastic Processes**, Holden–Day, 1962

## Appendix. FRACTAL MODELING OF SOIL MEDIA

The basic purpose of this section is to investigate whether soil microstructure characteristics can be modeled by fractal sets. The purely geometrical part of this work involves the setting up of a model that simulates the soil structure in appearance. It is necessary, however, to go beyond this and to devise a model that will not merely resemble the soil in appearance but also in mechanical behavior. By this is meant the deformation under loading, including the failure mechanism phenomenon.

In the following, a procedure is summarized for calculating the deformation under loading of a fractal set model. While the treatment is necessarily brief, there is sufficient detail devoted to particular simple examples to demonstrate the feasibility of a program that has the potential of initiating a subject that can become as voluminous as continuum mechanics.

It is the usual practice in studying the deformation of structural materials to model these materials as continua. By a continuum we mean a material whose mass and elastic (or inelastic) parameters vary continuously through the medium. In short, no voids of discontinuities are permitted in such a model, except a manageably small finite number of these. This permits the powerful methods of calculus to be employed, resulting in a system of differential equations, the solution of which leads to a complete description of the deformation of the material under an arbitrary loading.

Microscopic examination shows that no material can be described as a continuum. Thus, a continuum model, if it works at all, works when the deformations vary continuously when averaged over sufficiently large volumes. This can lead to some inadequacy in dealing with important inelastic phenomena, such as failure. It might be said then that continuum models are in their predominance today due to mathematical convenience rather than to physical observation.

Lately the fractal set has gained attention in that many structures that are found in nature are more readily modeled as fractal sets rather than continua. In this work, we intend to show how such a fractal set model can lead to a theory of deformation which parallels that of a continuum material model.

We will forgo generality and confine ourselves to the simplest possible example that displays the basic ideas. This will enable us to avoid the formidable technicalities which surround the theory of fractals. Thus, only an intuitive notion of a fractal set will suffice to understand our approach.

We now give a complete solution to one of the simplest problems imaginable: the fractal analog of the one–dimensional stress–strain test. Imagine a layered medium, which is composed of rigid layers interspersed with elastic layers. To convert the problem into one of fractal mechanics, we assume the medium is of unit depth and that the rigid layers occupy precisely the locations of the middle–third intervals.

To make this fractal model more plausible, one may consider it as the limit of a succession of mass–spring models. This is shown in Figure 3, where the first two stages are shown. In the first stage, shown in Figure 3a, the total mass $m$ is concentrated in a single block of length $1/3$, while two springs $k$ of equal length make up the rest of the model. Upon further resolution shown in Figure 3b, the springs further resolve into 4 springs of length $1/3^2$, and so on. In the limit, the springs will occupy no length at all! In contrast, consider the equivalent continuum model, which is conceived as being the limit of mass–spring models, as shown in Figure 4. In the limit the springs occupy the total length of the model, while the masses occupy zero length. Viewed in this manner, the fractal model is no more bizarre than the continuum model.

The equations for the continuum approximation illustrated in Figure 4b are given as follows:

$$\begin{aligned} \frac{m}{3}\ddot{u}_1 + k(2u_1 - u_2) &= F_1 \\ \frac{m}{3}\ddot{u}_2 + k(-u_1 + 2u_2 + u_3) &= F_2 \\ \frac{m}{3}\ddot{u}_3 + k(-u_2 + 2u_3) &= F_3 + k\Delta \end{aligned} \tag{1}$$

Let us consider the static case with $F_1 = F_2 = F_3 = 0$. Then the system becomes:

$$\begin{aligned} 2u_1 - u_2 &= 0 \\ -u_1 + 2u_2 + u_3 &= 0 \\ -u_2 + 2u_3 &= \Delta \end{aligned} \tag{2}$$

for which the solution is given by $u_1 = \frac{\Delta}{4}, u_2 = \frac{\Delta}{2}, u_3 = \frac{3}{4}\Delta$. If we rewrite this solution as $u_k = k\Delta/2^2$ for $k = 1,2,3$, then it will be apparent that at the $n^{th}$ stage, when there are $2^n - 1$ masses, the solution will be given by $u_k = k\Delta/2^n$ for $k = 1,2,\ldots,(2^n - 1)$. Since the masses are equidistant, tne solution approaches the linear function $u(x) = k\Delta$.

When the same approach is applied to the fractal approximations given in Figure 3b, we obtain exactly the same equations given in (2); (note, however that the masses in the dynamic version of (1) would be slightly different). Thus, the solution would be exactly the same. But the interpretation of $u_k$ is quite different; in the limit the solution is given by $u(x) = f(x)\Delta$, where $f$ is known as the Devil's Staircase function [17].

As a matter of fact, had we known nothing of the Devil's Staircase, we could have used this example to define it. Pushing this line of thought further, we can define other fractal functions. Refer again to the continuum model in Figure 4. Now let $\Delta = 0$ and all $F_i = F$. This case models a rod acted upon by its own weight. Proceeding as before, but omitting details, when $n \to \infty$, we find the expected solution, that $u(x)$ is proportional to the quadratic function $x(1-x)$. If we follow the same procedure in the case of the fractal model shown in Figure 3, we will obtain a fractal function which is the analog of the quadratic function. In like manner other fractal functions can be defined.

The first thing to note about the examples above is that both the continuum and fractal models can be approximated by systems having a finite number of degrees of freedom, and when they are so approximated, the resulting equations are identical. It is in the interpretation of the solution of these equations that the difference lies.

Consider the expression, $u_{i-1} - 2u_i + u_{i+1}$. For the continuum model, this expression becomes proportional to the second derivative of $u$ at some point on the model, as the degree of approximation goes to infinity. This is so because $u_i$ represents the motion at a point, and the relevant points are equidistant.

But in the fractal model, $u_i$ represents the deflection of an entire interval, and no such interpretation can be ascribed to $u_{i-1} - 2u_i + u_{i+1}$. Indeed, the limit function $u$ is not differentiable, and it is meaningless to speak of a fractal model being governed by a differential equation.

The second thing to note is that for the fractal model solution, the deformation function $u(x)$ is given by a fractal function, i.e., a function that varies only on a fractal set. These functions may appear strange at first, in that they are nondifferentiable; nevertheless, they are quite amenable when they lie under an integral sign.

This means that the finite element method may be used in conjunction with a fractal model much the same way as with continuum models. One need only recall how the finite element method works. Over some simple geometric element, say a triangle or tetrahedron, a restricted class of

deformations is allowed. This class of deformations will consist of polynomials, the coefficients of which are linear functions of the deformations of the vertices of the elements. For a fractal model, one simply replaces the polynomials by suitable fractal functions. In other words, the finite element method can be employed for fractal models in much the same way as for continuum models; polynomial functions are replaced by fractal functions.

The method presented above is very general. It remains to elaborate these in two and three dimensions. This will involve the determination of fractal functions and their integration over various finite elements. Comparisons with continuum models will be useful. These ideas can, in principle, be generalized to more complex problems.

## APPENDIX. THE MARKOV METHOD [23]

In classical physics, a basic role is played by the fundamental principle of scientific determinism: from the state of a physical system at the time $t_0$, one may deduce its state at a later instant $t$. As a consequence of this principle one obtains a basic method of analyzing physical systems: the state of a physical system at a given time $t_2$ may be deduced from a knowledge of its state at an earlier time $t_1$ and does not depend on the history of the system before time $t_1$. An example, differential equations.

For physical systems which obey probabilistic laws rather than deterministic laws, one may enunciate an analogous principle: the probability that the physical system will be in a given state at a given time $t_2$ may be deduced from a knowledge of its state at any earlier time $t_1$, and does not depend on the history of the system before time $t_1$. Stochastic processes which represent observations on physical systems satisfying this condition are called **Markov** processes.

A special kind of Markov process is a **Markov chain**; it may be defined as a stochastic process with an evolution which may be treated as a series of transitions between the **states** of the system. These states have the property that the probability law of the future evolution of the process, once it is in a given state, depends only on this state and not on how the process arrived at this state. The number of possible states is either finite or countably infinite.

A discrete parameter (time in this case) stochastic process $\{X(t), t = 0, 1, 2, ...\}$ or a continuous parameter stochastic process $\{X(t), t > 0\}$ is said to be a **Markov process** if, for any set of $n$ time points, $t_1 < t_2, \ldots, t_n$, the conditional distribution of $X(t_n)$, for given values of $X(t_1), \ldots, X(t_{n-1})$, depends only on $X(t_{n-1})$, the most recent known value; in mathematical notation, for any real numbers $x_1, \ldots, x_n$, representing possible states of the system,

$$P[X(t_n) \leq x_n | X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}] = P[X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}].$$

This equation may be read to mean that, given the "present" of the process, the "future" is independent of its "past."

Markov processes are classified according to (i) the nature of the index set of the process (discrete or continuous time), and (ii) the nature of the **state space** of the process (discrete or continuous state).

A real number $x$ is said to be a possible value, or **state**, of a stochastic process $X(t)$ if there exists a time $t$ such that the probability $P[x - h < X(t) < x + h]$ is positive for every $h > 0$. The set of possible values of a stochastic process is called its state space. The state space is called discrete if it contains a finite or countably infinite number of states. A state space which is not discrete is called continuous. A Markov process with discrete state space is called a **Markov chain**. Integers $\{0, 1, ...\}$ are used to represent the state space of such a chain.

A Markov process is described by a **transition probability function**, often denoted by $P(x, t_0; E, t)$ or $P(E, t|x, t_0)$, which represents the conditional probability that the state of the system will at time $t$ belong to the set $E$, given that at time $t_0 < t$ the system is in state $x$. The Markov process is said to have **stationary transition probabilities**, or to be **homogeneous** in time, if $P(x, t_0; E, t)$ depends on $t$ and $t_0$ only through the difference $(t - t_0)$. We initially consider homogeneous in time models.

The theory of Markov chains is initially presented with the assumption that the transition probabilities and matrices are known. This, of course, is never the case; the estimation of these transition probabilities is usually the most difficult and elusive part of the analysis.

In order to specify the probability law of a discrete parameter Markov chain $\{X_n\}$ it suffices to state for all times $n > m > 0$, and states $j$ and $k$, the probability mass function

$$p_j(n) = P[X_n = j]$$

and the conditional probability mass function

$$p_{j,k}(m,n) = P[X_n = k|X_m = j].$$

The function $p_{j,k}(m,n)$ is called the **transition probability function** of the Markov chain. The probability law of a Markov chain is determined by $p_j(n)$ and $p_{j,k}(m,n)$ since for all integers $q$, and any $q$ time points $n_1 < n_2 < \ldots < n_q$, and states $k_1, ..., k_q$,

$$P[X_{n_1} = k_1, \ldots, X_{n_q} = k_q] = p_{k_1}(n_1)p_{k_1,k_2}(n_1,n_2)\ldots p_{k(q-1),k_q}(n_{q-1},n_q).$$

A Markov chain is said to be **homogeneous** (or to be homogeneous in time or to have stationary transition probabilities) if $p_{j,k}(m,n)$ depends only on the difference $n-m$. We then call

$$p_{j,k}(n) = P[X_{n+t} = k|X_t = j] \quad \textit{for any integer } t \geq 0$$

the n–step **transition probability function** of the homogeneous Markov chain $\{X_n\}$. In words, $p_{j,k}(n)$ is the conditional probability that a homogeneous Markov chain now in state $j$ will move after $n$ (time) steps to state $k$. The one–step transition probabilities $p_{j,k}(1)$ are usually written simply $p_{j,k}$, i.e.,

$$p_{j,k} = P[X_{t+1} = k|X_t = j] \quad \textit{for any integer } t \geq 0.$$

Similarly, if $\{X(t), t \geq 0\}$ is a continuous parameter Markov chain, then to specify the probability law of $\{X(t), t \geq 0\}$, it suffices to state for all times $t \geq s \geq 0$, and states $j$ and $k$, the probability mass function

$$p_k(t) = P[X(t) = k]$$

and the conditional probability mass function

$$p_{j,k}(s,t) = P[X(t) = k|X(s) = j].$$

The function $p_{j,k}(s,t)$ is called the **transition probability function** of the Markov chain. The Markov chain is said to be homogeneous (or to have stationary transition probabilities) if $p_{j,k}(s,t)$ depends only on the difference $t-s$. We then call

$$p_{j,k}(t) = P[X(t+u) = k|X(u) = j] \textit{ for any } u \geq 0$$

the n–step transition probability function of the Markov chain $X(t), t > 0$.

A fundamental relation satisfied by the transition probability function of a Markov chain $\{X_n\}$ is the so–called **Chapman–Kolmogorov** equation: for any times $n > u > m > 0$ and states $j$ and $k$,

$$p_{j,k}(m,n) = \sum p_{j,i}(m,u)p_{i,k}(u,n),$$

summing over all states $i$ of the Markov chain. This equation represents all possible intermediate states $i$ between $j$ and $k$.

The transition probabilities of a Markov chain $\{X_n\}$ with state space $\{0,1,2,...\}$ are best exhibited in the form of a matrix:

$$P(m,n) = \begin{bmatrix} p_{0,0}(m,n) & p_{0,1}(m,n) & \cdots & p_{0,k}(m,n) & \cdots \\ p_{1,0}(m,n) & p_{1,1}(m,n) & \cdots & p_{1,k}(m,n) & \cdots \\ & & \cdots & & \cdots \\ p_{j,0}(m,n) & p_{j,1}(m,n) & \cdots & p_{j,k}(m,n) & \cdots \\ & & \cdots & & p_{J,K}(m,n) \end{bmatrix}.$$

Note that the elements of a transition probability matrix $P(m,n)$ satisfy the conditions

$$p_{j,k}(m,n) \geq 0 \quad \text{for all } j,k$$
$$\sum_k p_{j,k}(m,n) = 1 \quad \text{for all } j.$$

The Chapman–Kolmogorov equations for all times $n > u > m \geq 0$ may be written:

$$P(m,n) = P(m,u)P(u,n).$$

From the Chapman–Kolmogorov equation, one may derive the recursive relation:

$$p(n) = p(0)P(0,n),$$

where

$$p(n) = [p_0(n)\ p_1(n) \ldots p_j(n) \ldots],$$
$$p_j(n) = P[X_n = j].$$

It follows that the probability law of a Markov chain $\{X_n\}$ is completely determined once one knows the transition probability matrices, and the unconditional probability vector $p(0)$ at time 0. In the case of a homogeneous Markov chain $\{X_n\}$, let

$$P(n) = \{p_{j,k}(n)\}, \quad P = \{p_{j,k}\}$$

denote respectively the $n$–step and the one–step transition probability matrices. From the above equations, it is observed that

$$P(n) = P^n,$$
$$p(n) = p(0)P^n.$$

Consequently, the probability law of a **homogeneous** Markov chain is completely determined once one knows the one–step transition probability matrix $P = \{p_{j,k}\}$, and the unconditional probability vector $p(0) = \{p_j(0)\}$ at time 0.

A Markov chain is said to be a **finite Markov chain** with $K$ states if the number of possible values of the random variables $\{X_n\}$ is finite and equal to $K$. The transition probabilities $p_{j,k}$ are then non–zero for only a finite number of values of $j$ and $k$, and the transition probability matrix $P$ is then a $K$x$K$ matrix.

A state $k$ is said to be **accessible** from a state $j$ if, for some integer $N \geq 1$, $p_{j,k}(N) > 0$. Two states $j$ and $k$ are said to **communicate** if $j$ is accessible from $k$, and $k$ is accessible from $j$.

A state $k$ is said to be **recurrent** if the probability is 1 that the Markov chain will eventually return to $k$, having started at $k$. A state $k$ is said to be **non–recurrent** if the above probability is less than 1.

A state $k$ is called an **absorbing** state if $p_{k,k} = 1$, so that once the chain visits $k$ it remains there forever. An absorbing state is clearly recurrent.

## APPENDIX. FIGURES

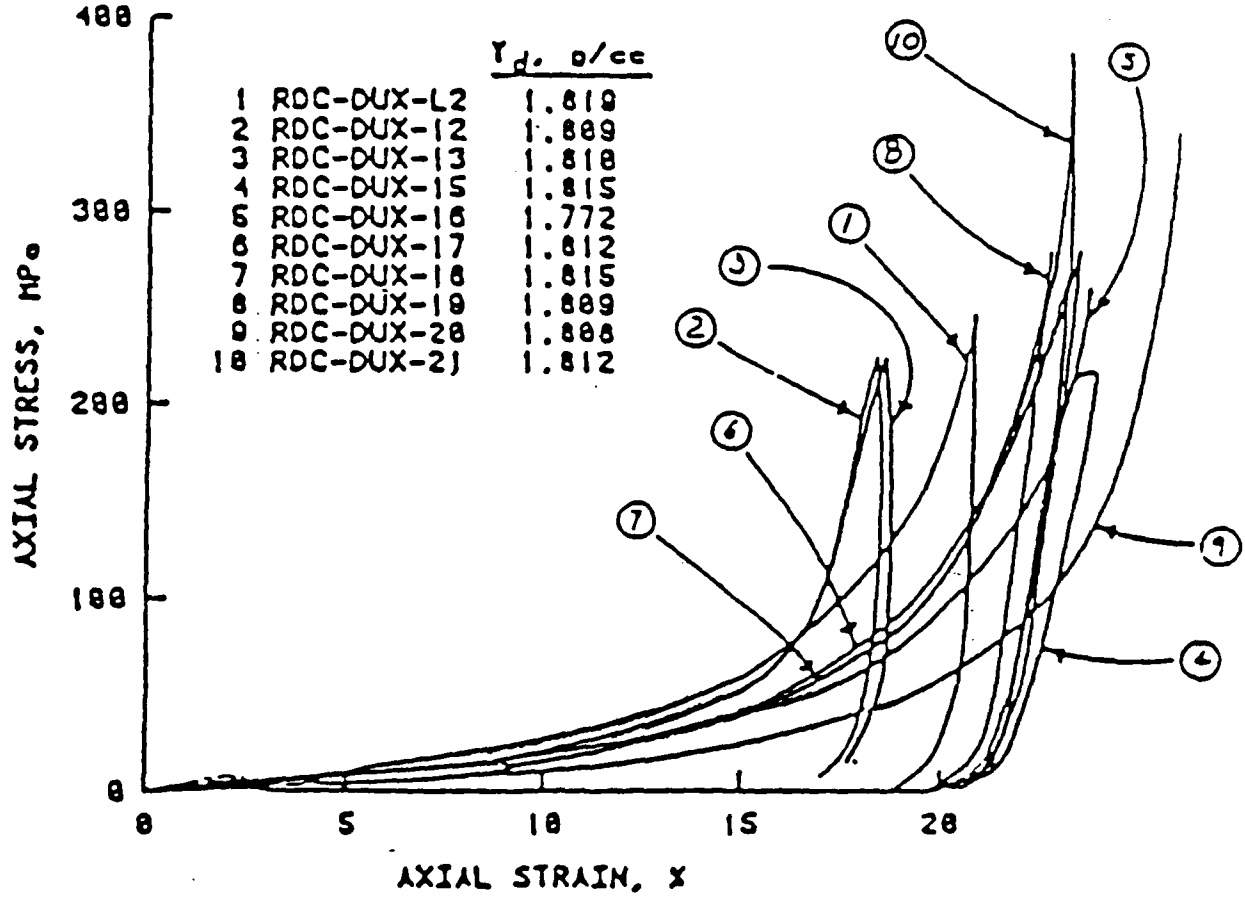


Figure 1. WES Data [7]

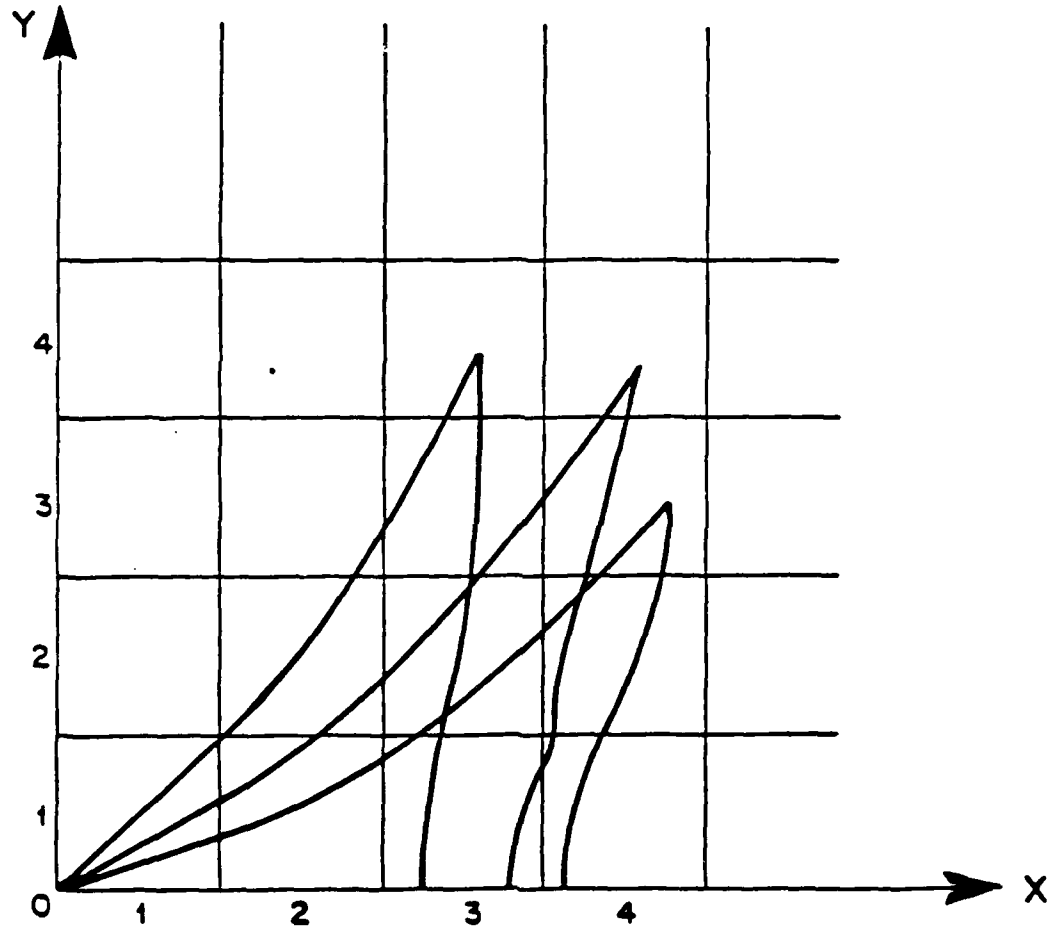


Figure 2. Generic Curves on Square Grid State Space

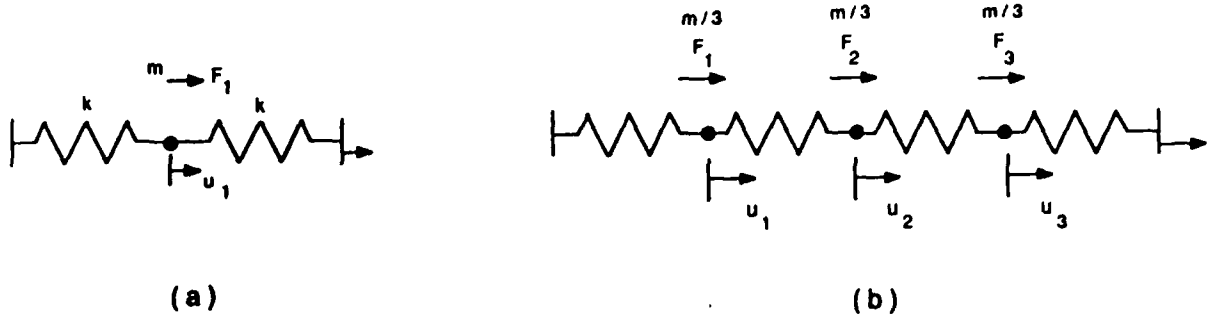


Figure 3. Fractal Model

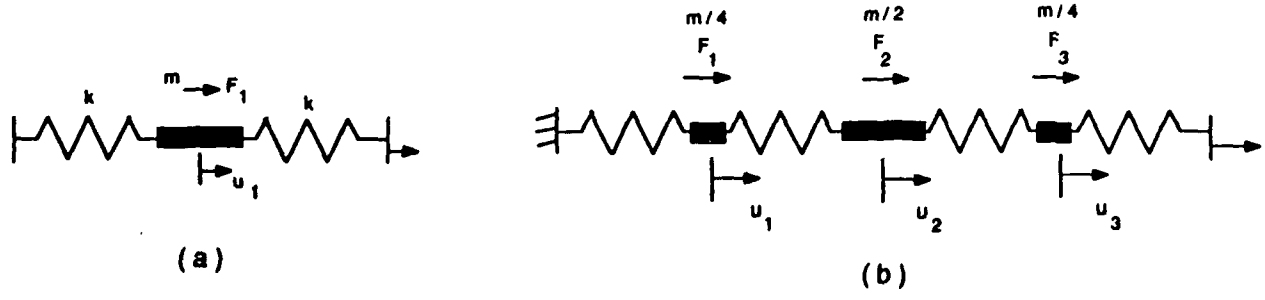


Figure 4. Continuum Model